



# **Fulcrum's FocalPoint II, A 300nS, 240 Gb/s switch/router**



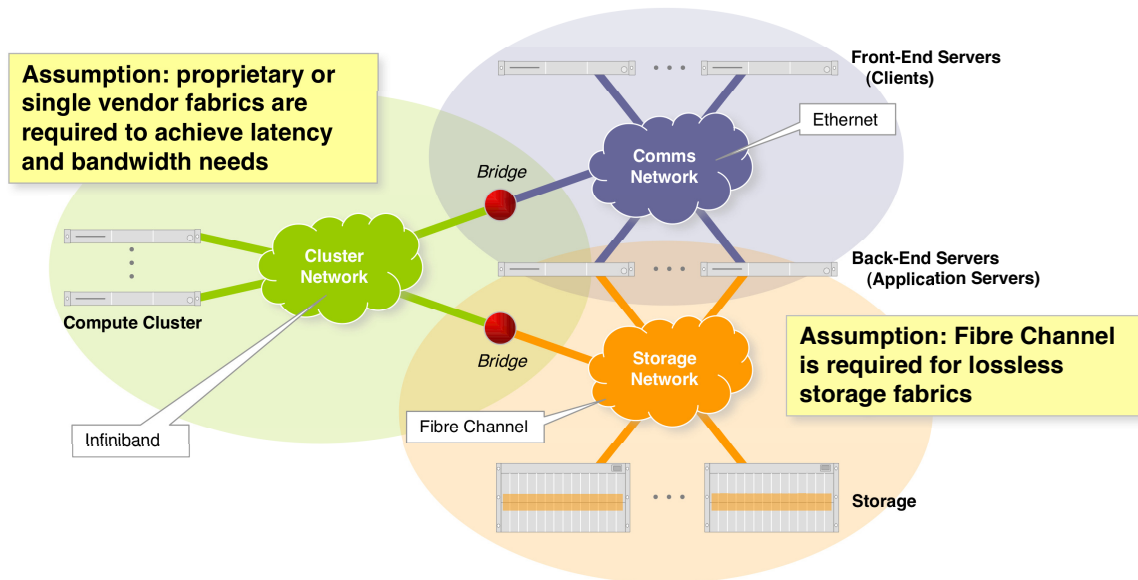
## **Agenda**

**Datacenter Interconnect Requirements  
and FocalPoint  
FocalPoint II (Bali) Chip**



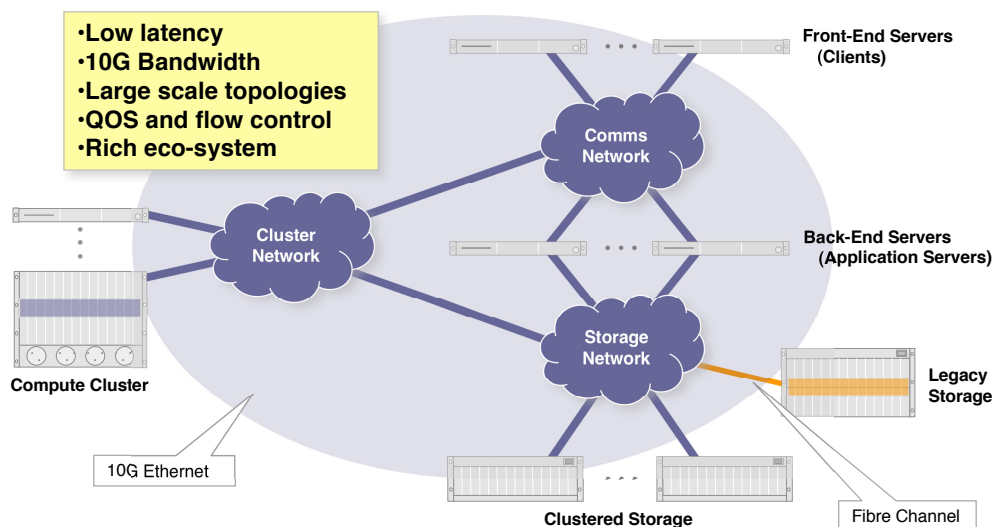
# Anatomy of the Multi-fabric Data Center

*Inefficient islands add complexity and cost; limit scale-out*



## 10GE is the unifying datacenter Network Technology

*Datacenter Ethernet enables full cross-sectional bandwidth and a single management domain over all three networks*

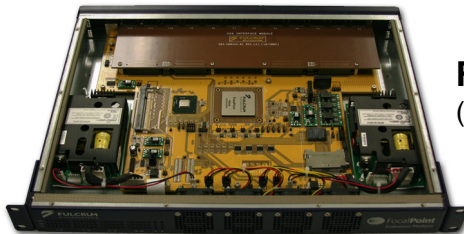


# Step 1: Solve Latency and Port Density

*The world's most powerful Ethernet switch chip*



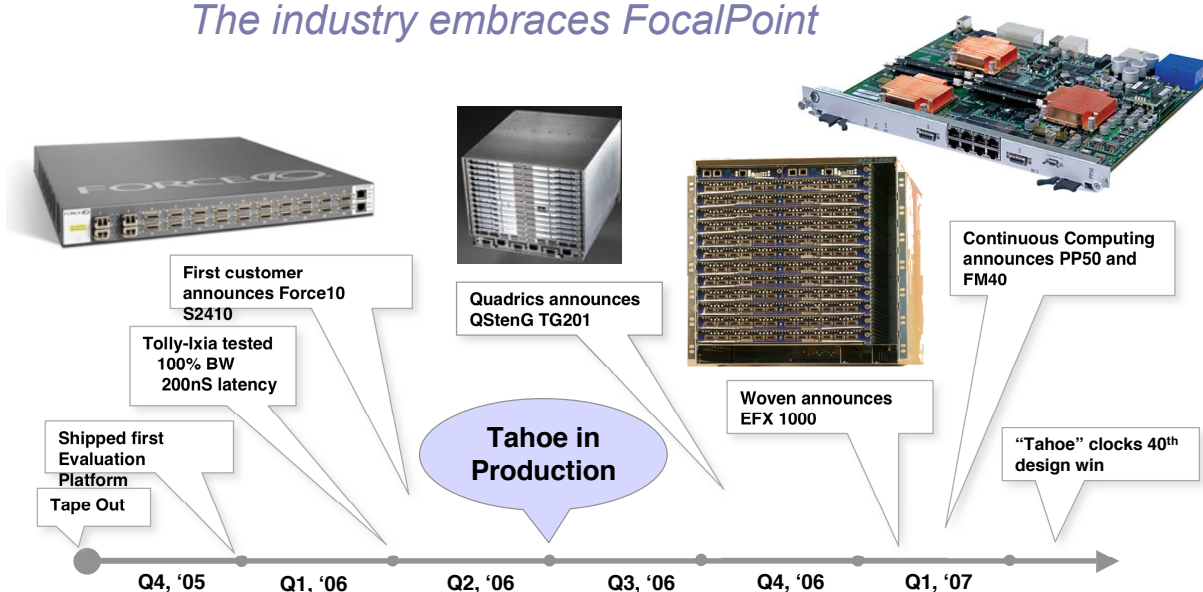
- **Highest port density** (24 10GE ports)
- **Lowest latency** (200ns)
- **Highest performance** (240Gbps)
- **Most power efficient** (<150mW/Gbps)
- **Most integrated** (single chip)
- **Most scalable** (fat trees, 1,000s of ports)



**FocalPoint Evaluation Platform**  
(The world's most integrated 10G Ethernet system)

## FocalPoint Status Report

*The industry embraces FocalPoint*



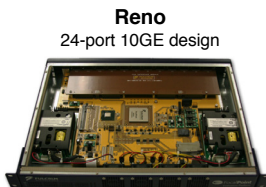
# Fulcrum and Partner Developed Platforms

- **System Reference Design**

- Reno (24-port 10G pizza box)
  - ControlPoint software golden model
  - Partner collaboration platform

- **Hardware Reference Designs**

- Lassen (24-port SFP+ pizza box, co-developed with AMCC)
- San Marino (10GBASE-T box, co-developed with Solarflare)
- Heavenly (IBM BC-H Fabric Card)
- Vegas (48 + 8 pizza box)



Fulcrum Confidential



7

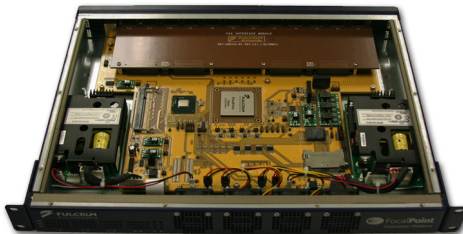
## Agenda

**Datacenter Interconnect Requirements  
and FocalPoint  
FocalPoint II (Bali) Chip**



## Step 2: Routing & Network Performance

### *FocalPoint II (Bali) project goals*



- **Maintains Gen I Performance**
    - 24 10GE ports, 200nS, 360 MPPS
    - But increase to 2MB memory
  - **IPv4 & IPv6 unicast & multicast routing**
    - 16k IP addresses
  - **L2-L4+ ACLs with deep inspection**
  - **Chip cascades**
    - Virtual switch of Clos, rings & meshes
  - **Clos Improvements**
    - As close to full bandwidth as possible
- Converged Enhanced Ethernet (CEE)**
- Enable lossless Ethernet fabrics

www.fulcrummicro.com

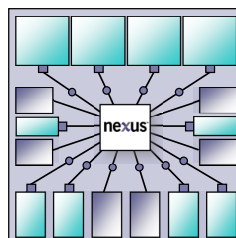
9



## Architecture Enabling Circuits

### *Two key IP blocks demonstrate the virtues of the technology*

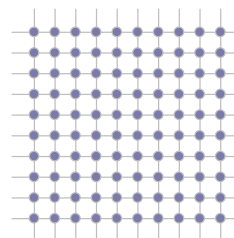
**Nexus\***  
(Terabit Crossbar)



- Gigahertz performance
- Terabit capacity
- Nanosecond latency
- No power penalty

\* Licensed to **PMC** for SoC interconnect  
**PMC-SIERRA**

**RapidArray**  
(Asynchronous SRAM)



- 600 MHz SRAM
- 1200 MHz interconnect
- 518 Gbps throughput
- Scalable for any use

### Key Benefits:

- Easily integrates independent clock domains
- Provides 4x overspeed
- Reduces overall chip area
- 2x the speed of vendor cores (same size, density, yield)
- Reduces power consumption (based on activity)
- CAM circuit is close relative

www.fulcrummicro.com

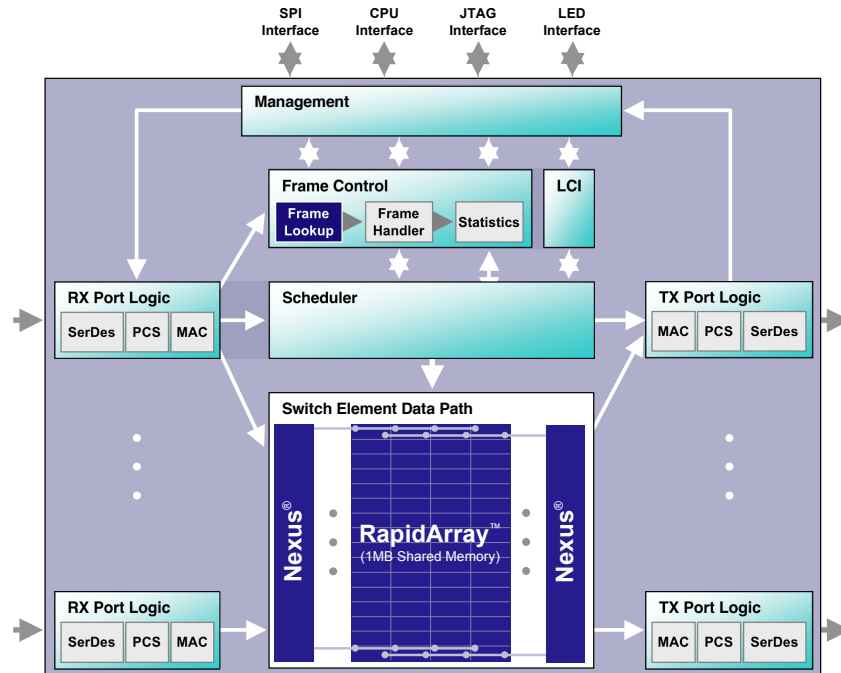
10





# FocalPoint I & II Architecture

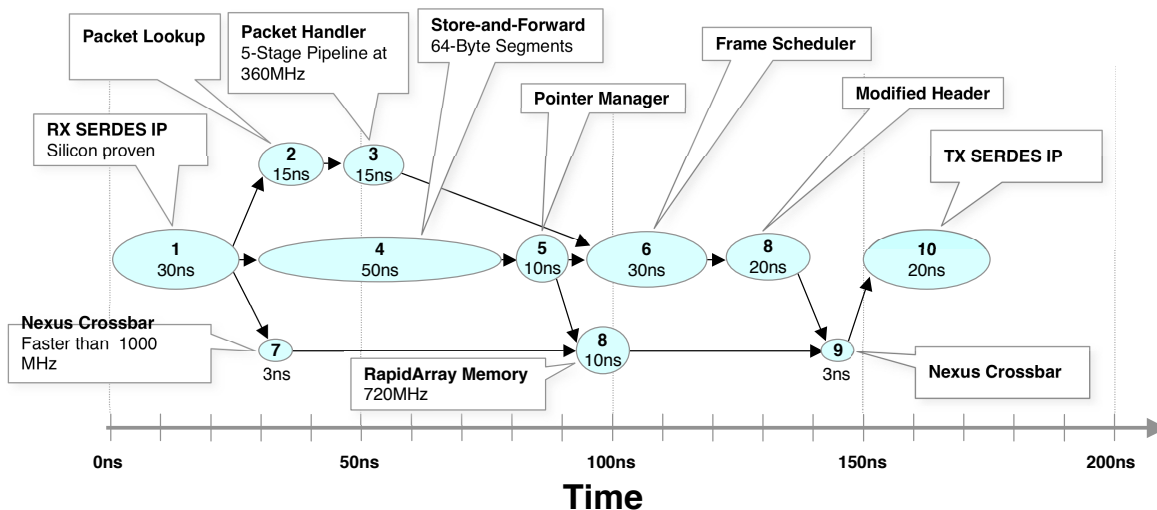
*Modular architecture, centralized control*



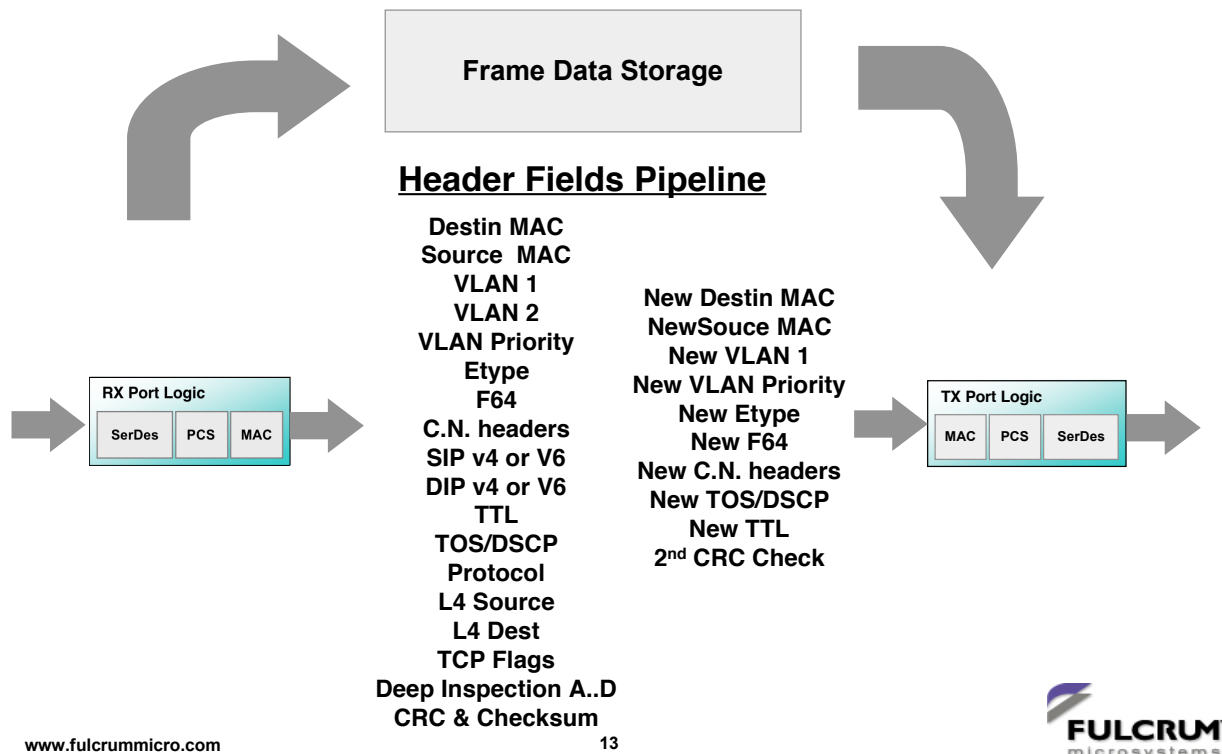
# FocalPoint Gen I & II Latency Detail

*Requirement: Full L3 latency less than 300 ns with 360 MPPS*

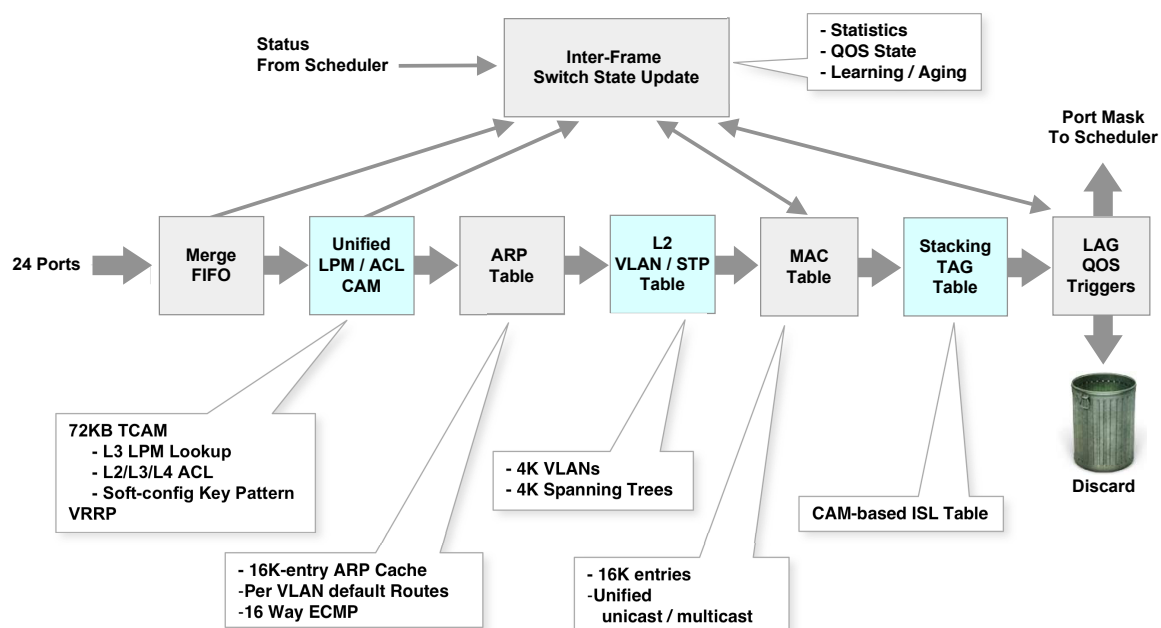
*Budget: 35 clock cycle pipeline (2.6 nS / cycle) for all L3*



# L2-L4 Packet parsing and manipulation

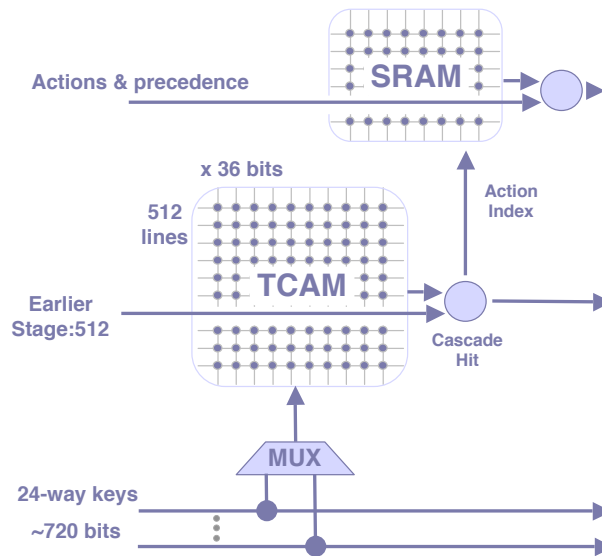


# FPII L2-L4 Packet Processing Pipeline

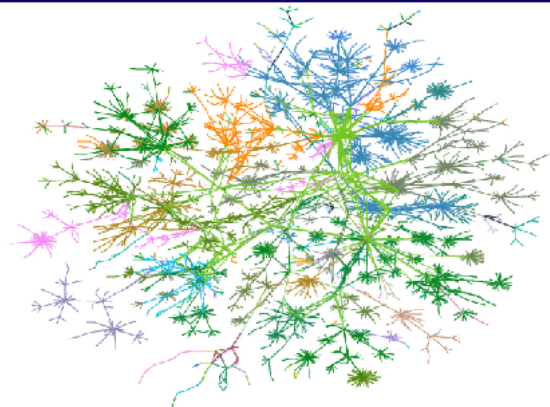
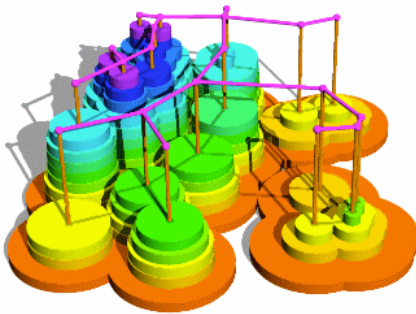


# Filter & Forwarding (FFU) Organization

- **TCAM Organization**
  - 32 slices
  - Cascades
  - Precedence
- **Performance**
  - 1 delay per stage
  - 6 stages per flop
  - 6 clocks overall
- **Flexibility**
  - Any key available
- **Compression**
  - Mapper unit (not shown)
  - Derived keys for ranges and repeated patterns



# Topology Enhancements and Ethernet

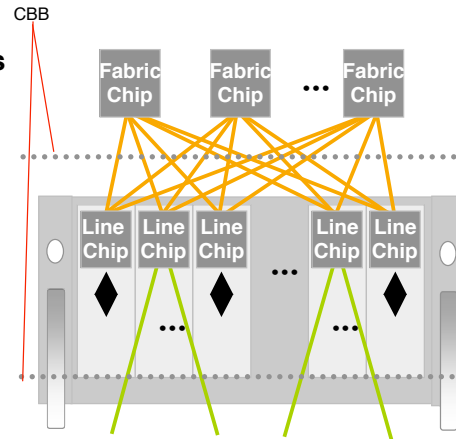


- **A Chip is not an Ethernet switch**
  - We may want to use one or many chips in a “switch”
  - We need to link data plane port state between chips
  - Multi-chip and multi-box link aggregation and management
- **An Ethernet switch uses a single spanning tree**
  - Networks should scale in a non-blocking fashion
  - Spanning tree hashing and resilience



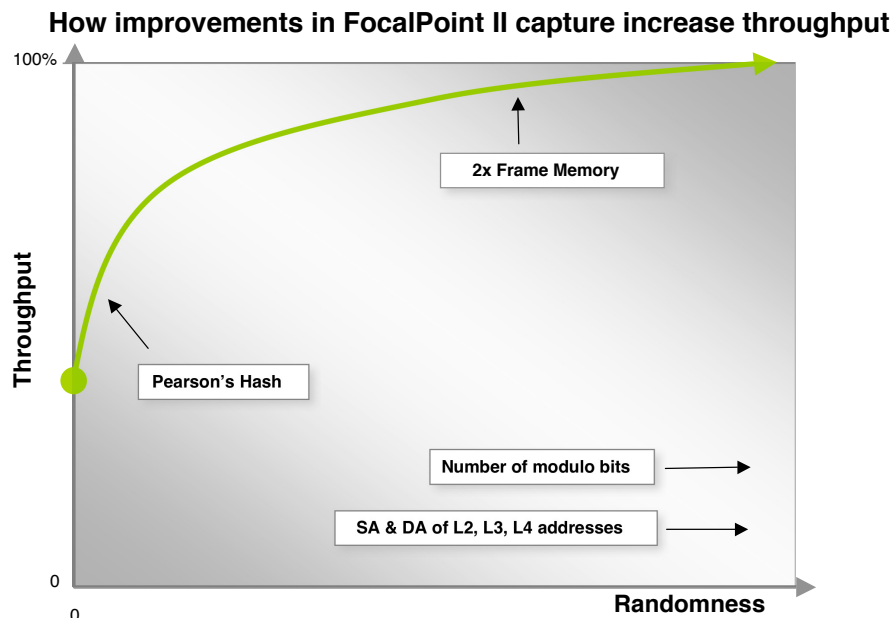
# Clos Architecture Theoretical Performance

- **Economic advantages are driving the move to Clos**
  - 10x reduction in per port price
  - 10x reduction in latency
  - 2-8x increase in port density / BW
- **All modern Clos architectures are statistical multi-path, multi-hop networks**
  - Examples are Infiniband (Mellanox), fiberchannel (Brocade), Ethernet (Fulcrum)
- **Clos architectures achieve ideal performance if**
  - The path selection is sufficiently stochastic
  - There is enough over-speed in the system to compensate for any non-ideal path selection
  - There is enough memory per switch to compensate for collisions so that flow control is infrequent



## FocalPoint II achieves new performance levels

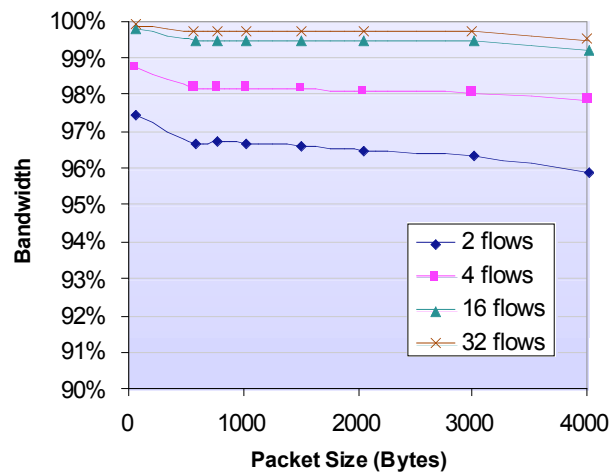
*All flow based systems face the same challenges*



# Simulated Clos Performance

- **Standard Full Mesh**
  - Each port sends to every other port in the system every cycle
  - Every packet is randomly assigned to a flow on its port
  - 2-32 flows per port simulated
  - 576-9216 system flows (288P)
  - 288P system is made from 36 24P switch chips in a Clos configuration
- **Amount of over-speed from line card switching (24P line cards)**
  - 3% in 288P Clos
  - 23% in 48P Clos
- **Conclusion**
  - 16 flows per port is nearly perfect in a 288P system (four flow is very good)
  - 1 flow per port is nearly perfect in a 48P system given the 23% overspeed

288P Full Mesh Performance

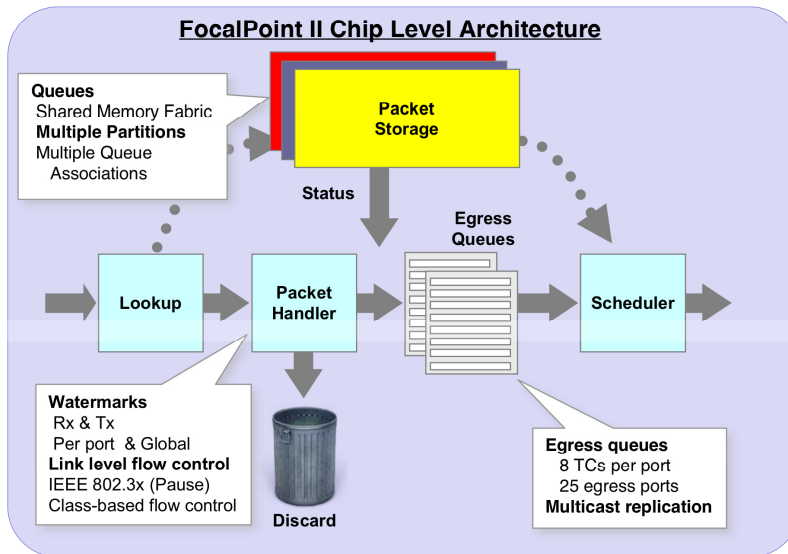


# Congestion Control

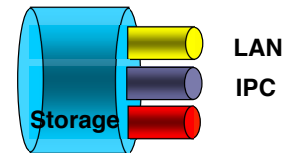
- **There are three drivers for congestion control enhancements in the enterprise datacenter**
  - Storage & IPC traffic is highly loss and jitter sensitive
  - Multiple traffic types, like storage and LAN, require different best-in-class congestion control practices
  - Cost oriented fully integrated full bandwidth switch chips are required to use memory very efficiently
- **As a result there is a race to produce lossless, non-HOL blocking, low latency fabrics with optimal bandwidth**
  - Congestion control is being standardized by the IEEE in 802.1au and potential future working groups

# Congestion Management and Scheduling Architecture

*Traffic separation enables virtual switching*

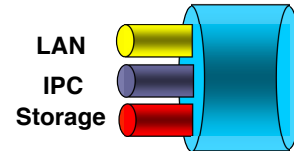


## Egress Features



- Multi-level scheduling  
Bandwidth groups with priority
- Traffic distribution  
Deficit weighted round robin

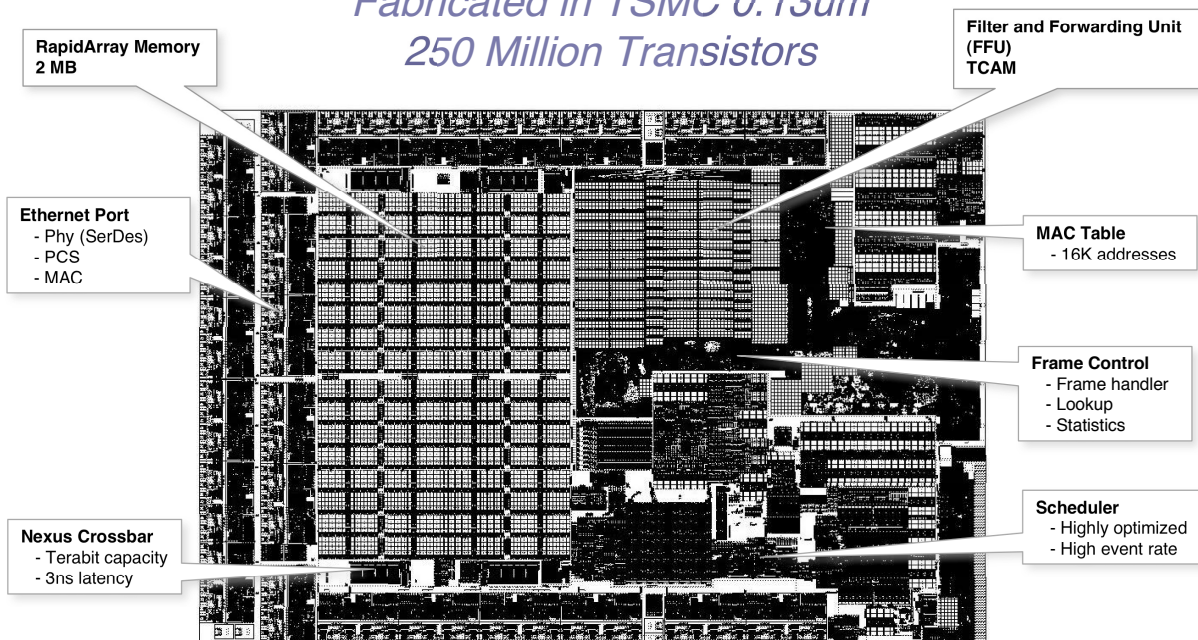
## Ingress Features



- Flow Control  
Link & Per class pause
- Static & Dynamic rate-limiting  
Pause pacing & policing  
NIC Proxy

# Bali Chip Plot

*Fabricated in TSMC 0.13um  
250 Million Transistors*



# Thank You!

**Uri Cummings**

*Founder, CTO*

uri@fulcrummicro.com



818.871.8100  
www.fulcrummicro.com

26630 Agoura Road  
Calabasas, CA 91302